



ZPRÁVA ZE SEMINÁŘE

Uživatelské statistiky EIZ

20.5.2015

Marta Zizienová

Místo schůzky : Národní technická knihovna

Seminář probíhal dle programu. Z pohledu UKN je zajímavé:

1. Měření využívanosti EIZ, standard COUNTER – Jiří Jirát.

Zjednodušeně a nesprávně jsou nazývané „uživatelské statistiky“. Měříme „stažení článku“ v rámci webových služeb. U článku v HTML je několik příkazů GET, u PDF je to většinou 1 GET. U https nejsou jiné možnosti než získat data od vydavatele, není možné se podívat na server a zjistit kvantitu komunikace se serverem (není možné se podívat do provozu).

Je otázka, kde budeme data v rámci sítě měřit.

Vydavatel měří buď analýzu logů webserveru (může vést k podhodnocení statistik, např. u stránek v cache); page tagging (nutné JavaScript, cookies, vyžaduje spolupráci uživatele; výrazně více šmíruje uživatele).

Vlastní statistiky – můžeme sledovat provoz na firewallu (asi nejúplnější; šlo by dohledat až na konkrétního uživatele; otázka zákonnosti; politicky nevhodné, data jsou hrubá a obrovská); logy z webové proxy (úplnost záleží na politice instituce a uspořádání sítě; lze dohledat na úrovni uživatele; pravděpodobně politicky neprůchodné; nutný filtr, velká množství dat); provoz na EZproxy (jsou data jen k EIZ a ne z jiného provozu; v logu jsou i citlivá data; nevýhoda – malé celkové využití – cca 3-5%); nelze odlišit placený/free/OA obsah); linking server (statistiky jsou už připravené, data jsou bez balastu, malý provoz 3-4%; cenným údajem je to, odkud uživatelé chodí). Data od vydavatelů a data od samotných knihoven se někde liší (ne že by vydavatelé lhali, ale data jsou různá).

Dalšími metrikami jsou abstract views nebo searches. V prezentaci je tabulka srovnání jednotlivých metod. Nejlépe vycházejí data od vydavatele nebo poskytovatele, ta by měla pokrývat 100% provozu a je to to nejlepší, žádná jiná alternativa není dostatečně objemná.

Standard COUNTER standardizuje způsob nahlašování provozu. Platná je release 4. Vydavatel musí podstoupit roční nezávislý audit, na webu je seznam COUNTER-compliant poskytovatelů.

Terminologie: Automated search = hledání z discovery vrstvy. Federated search = hledání ve více databázích. Regular search = čistá činnost člověka. Session = úspěšné využití služby. Result click = kliknutí na výsledek, record views = počet zobrazení záznamů/abstraktů. Turnaways nebo Denials = odmítnuté přístupy. Access denied = nemám zdroje předplacené nebo není dost licencí.

Reporty pro konsorcia – dodavatel má povinnost dodat agregovaný konsorciální report a reporty pro jednotlivé členy konsorcia.

U časopisů jsou dostupné Journal Report 1 (počet úspěšně stažených plných textů z placených i neplacených zdrojů dle jednotlivých měsíců); Journal Report 1 GOA (podmnožina JR1 - počet přístupů v rámci gold OA); Journal Report 1a (není povinný; počet z archivu/backfile); Journal Report 2 (kam uživatelé chtěli, ale nedostali se); Journal Report 5 (ukazuje, z jakých roků uživatelé čte; obsahuje data za aktuální a předchozí dekádu).

U databází je k dispozici Database Report 1 (jak uživatelé hledali, klikali a zobrazovali záznamy); Database Report 2 (odmítnuté přístupy); Platform Report 1 (využívání platformy s více



databázemi).

U knih jsou Book Report 1 (využívání knih, které JSOU JAKO CELEK, nejsou rozsekané na kapitoly); Book Report 2 (u knih, kde si uživatelé zobrazují jednotlivé sekce knihy; obsahuje i typ využití sekce); Book Report 3 (kam se uživatelé chtěli podívat a nedostali se tam); Book Report 4 (počty odmítnutí na úrovni platformy); Book Report 5 (jak uživatelé hledají v konkrétním titulu, jen u knih na úrovni titulu, u knih, které mají vlastní vyhledávání – např. encyklopedie).

Existuje ještě Multimedia Report 1 (jen u databází, kde existuje multimediální obsah – samostatné obrázky, videa, zvuk).

Pokud vydavatel vůbec neposkytuje statistiky, je to často malý vydavatel. Pokud statistiky nemám, bojím se předplatné zrušit a vyjednávat o ceně.

2. Automatizované stahování dat o využívanosti EIZ – protokol SUSHI – Jan Pokorný

Jedná se o normu NISO Z39.93:2014, webovou SOAP službu client/server na portu 80, která v pravidelných intervalech stahuje data ve standardizovaném formátu. Klientem je typicky ERM systém nebo jiný nástroj. Pomůcky pro automatické stahování jsou:

www.niso.org/workrooms/sushi (seznam provozovatelů, kteří jsou certifikováni, jaké reporty poskytují), SUSHIStarters Client (online možnost ověření, lze získat konkrétní report). V ERM systému vidím, jak probíhaly konkrétní sklízňe a co do ERM natéká (většinou je také možnost přidat statistiku ručně).

Pozor, v ERMS se často přehled nazývá report (což není totéž jako report COUNTERu).

K dispozici je pycounter (2 moduly pro python, umožňují sklizení přes SUSHI a práci se statistikami COUNTER). V prezentaci je příklad kódu.

ERM systémy se SUSHI jsou EBSCO Usage Consolidation, 360 Counter, Alma, open source - Coral ERMS, ERMes, Quali OLE.

Typické problémy jsou – občasné výpadky webových služeb, podpora různých verzí COUNTERu a různých typů reportů, různá interpretace ukazatelů aj.

Americké univerzity vytvářejí bezplatnou knowledge base (GOKB), která je volná (už nechtějí platit vysoké částky za komerční knowledge base). Ideální mít ERMS a discovery od jedné firmy, aby systémy byly kompatibilní a data úplná.

3. Technické aspekty ovlivňující využívanost zdroje – Jiří Jiráček

Čím snadněji se uživatel dostane na plný text, tím vyšší je statistika stažení plného textu (např. při přístupu přes CrossRef; v té chvíli je otázkou, jestli doopravdy chtěl plný text). Většina uživatelů čte html verzi, ale tiskne/stahuje PDF, TOTAL se překrývá. Drasticky statistiky ovlivní federované vyhledávače, crawlers, roboti, nástroje pro hromadné stahování, referenční manažery (ReadCube, Mendeley ...). Otázkou je prefetching u discovery systémů nebo prohlížečů, double-clicking. Uživatelské rozhraní má měřitelný vliv na využívanost (ideálně je jeden klik mezi citací a plným textem, vyhledatelnost). Velký vydavatel se snaží optimalizovat data, aby bylo vidět, že databáze je využívána.

Z měření se dá vyčíst, jaké máš čtenáře, podpůrné argumenty pro udržení/zrušení zdroje, vytipování nových akvizic, ale je třeba se bavit s uživateli. Zásadní je interpretace – jestli to, co uživatelé naklikali, bylo užitečné, to nejde zjistit. Nezbytné je víceleté srovnání (v prezentaci jsou základní argumenty), důležitý je kontext.

Zpracovala: Marta Zizienová

